Künstl Intell (2010) 24: 345–348
DOI 10.1007/s13218-010-0059-6

DISSERTATIONEN UND HABILITATIONEN

Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments

Radu Bogdan Rusu

Published online: 17 August 2010

Abstract Environment models serve as important resources for an autonomous robot by providing it with the necessary task-relevant information about its habitat. Their use enables robots to perform their tasks more reliably, flexibly, and efficiently. As autonomous robotic platforms get more sophisticated manipulation capabilities, they also need more expressive and comprehensive environment models: for manipulation purposes their models have to include the objects present in the world, together with their position, form, and other aspects, as well as an interpretation of these objects with respect to the robot tasks.

The dissertation presented in this article (Rusu, PhD thesis, 2009) proposes *Semantic 3D Object Models* as a novel representation of the robot's operating environment that satisfies these requirements and shows how these models can be automatically acquired from dense 3D range data.

1 Why 3D Semantic Perception?

Humans perceive their environments in terms of images, and describe the world in terms of what they see. This erroneously suggests that we might be able to frame the perception problem solely in a 2D context. As we can see from Figs. 1 and 2, framing the perception problem this way can lead to failures in capturing the true semantic meaning of the world.

The reasons are twofold. On one hand, monocular computer vision applications are flustered by both fundamental deficiencies in the current generation of camera devices and limitations in the datastream itself. The former will most likely be addressed in time, as technology progresses and better camera sensors are developed. An example of such a deficiency is shown in Fig. 1, where due to the low dynamic range of the camera sensor, the right part of the image is completely underexposed. This makes it very hard for 2D image processing applications to recover the necessary information for recognizing objects in such scenes.

Fig. 1 Model matching failure in underexposed 2D images. None of the features extracted from the model (*left*) can be successfully matched onto the object of interest (*right*)

Fig. 2 An example of a good model match using features extracted from 2D images (*left*), where an object template is successfully identified in an image. However, zooming out from the image, we observe that the template is in fact another picture stitched to a completely different 3D object (in this case a mug). The semantics of the objects in this case are thus completely wrong

R.B. Rusu (✉)
Intelligent Autonomous Systems, Computer Science Department, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany
e-mail: rusu@cs.tum.edu
url: http://www.rbrusu.com

The second reason is linked directly to the fact that computer vision applications make use of 2D camera images mostly, which are inherently capturing only a projection of the 3D world. Figure 2 attempts to capture this problem by showing two examples of matching a certain object template in 2D camera images. While the system apparently matched the model template (a beer bottle in this case) successfully in the left part of the image, after zooming out, we observe that the bottle of beer was in fact another picture in itself, stitched to a completely different 3D geometric surface (the body of a mug in this case). This is a clear example which shows that the semantics of a particular solution obtained only using 2D image processing can be lost if the geometry of the object is not incorporated in the reasoning process.

2 The Computational Problem

The main computational problems of semantic perception that are investigated in this thesis are depicted in Figs. 3 and 4. In both figures, the input data consists of range measurements coming from 3D sensing devices, thus the input is a set of points with each point having a 3D coordinate in the world.

The first computational problem that we want to investigate can be formulated as follows. Given a set of individual datasets as the ones presented in the left part of Fig. 3, build a Semantic 3D Object Map such as the one presented in the right part of the figure, that contains information about the objects in the world such as: cupboards are cuboid containers with a frontal door and a handle, tables are supporting planes which can hold objects that are to be manipulated, and kitchen appliances contain knobs and handles that can be used to operate them [5].

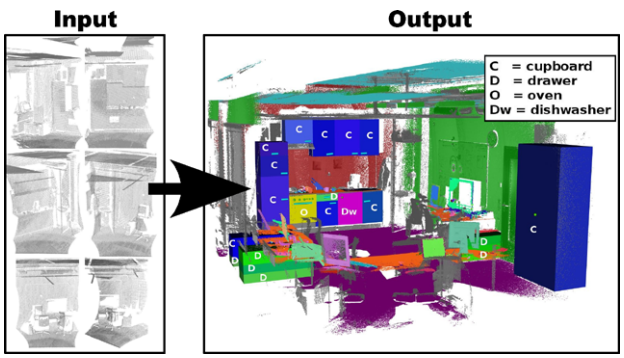

Fig. 3 Computational problem 1: create a Semantic 3D Object Map representation (*right*) for a kitchen environment from a set of input datasets (*left*)

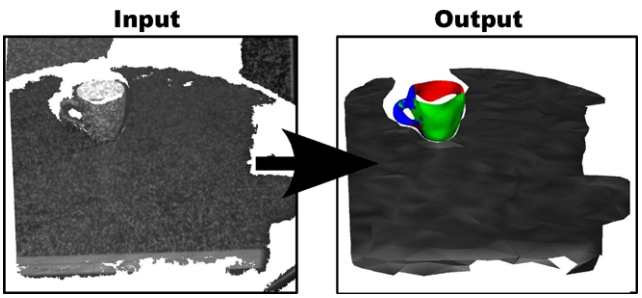

Fig. 4 Computational problem 2: decompose an object into individual parts and reconstruct its surface (*right*) from a raw dataset (*left*)

The second computational problem is given in Fig. 4. Given a dataset representing a supporting table plane with an object present on it (left), produce a representation such as the one shown on the right part, where the object model is decomposed into individual parts to help grasping applications and its surface is reconstructed using smooth triangular meshes [4].

These two examples pose serious challenges for the current generation of 3D perception systems. Though both applications are formulated in the context of structured human living environments, the variation of object placements and shapes, as well as the differences from one room and apartment to another, makes the problem to be solved extremely difficult. In particular the acquired datasets contains exacerbated levels of noise, there are objects and surfaces which do not return measurements thus leaving holes in the data, the scenes are too cluttered and the objects to be searched for contain extremely large—if not infinite—variations of shapes and colors, just to mention a few.

3 Semantic 3D Object Maps

The contributions of this thesis consist in numerous theoretical formulations and practical solutions to the problem of 3D semantic mapping with mobile robots. Divided into three distinct parts, the thesis starts by presenting important contributions to a set of algorithmic steps comprising the kernel of a Semantic 3D Object Mapping system [2]. This includes dedicated conceptual architectures and their software implementations for 3D mapping systems, useful in solving complex perception problems. Starting from point-based representations of the world as preliminary data sources, an innovative solution for the characterization of the surface geometry around a point is given through the formulation of a Point Feature Histogram (PFH) 3D feature [3]. The proposed feature space is extremely useful for the problem of finding point correspondences in multiple overlapping datasets (see Fig. 5), as well as for learning classes of geometric primitives that label point sampled surfaces (see Fig. 6). Other important contributions include partial view point cloud registration under geometric constraints, and surface segmentation and reconstruction from point cloud data for static environments.

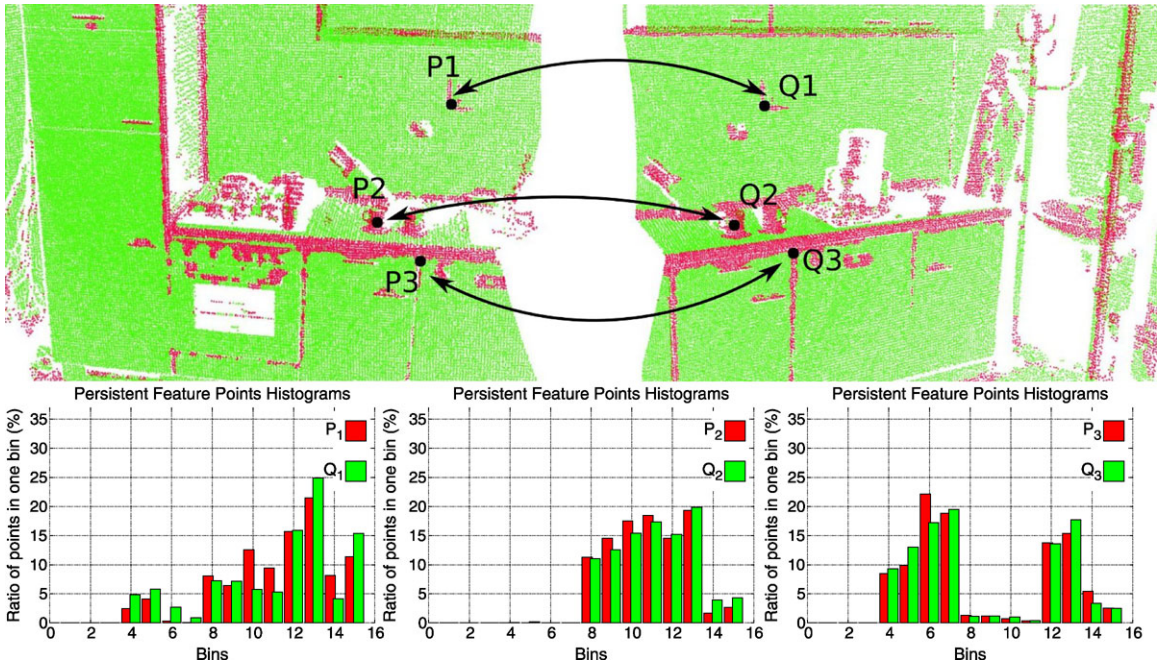

Fig. 5 Point feature representations for three pairs of corresponding points (p_i, q_i) on 2 different point cloud datasets

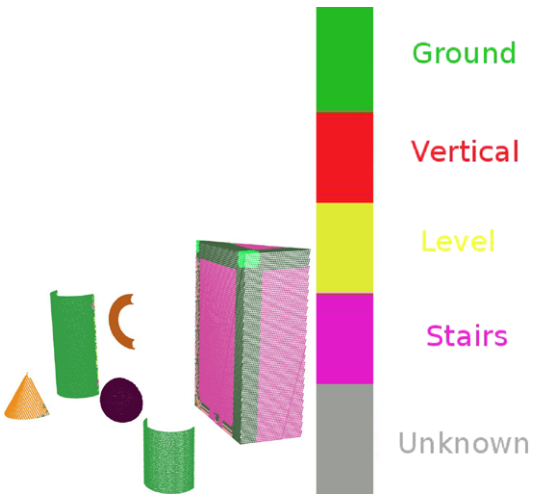

Fig. 6 Classification results using Point Feature Histograms for a synthetic scene using Support Vector Machines

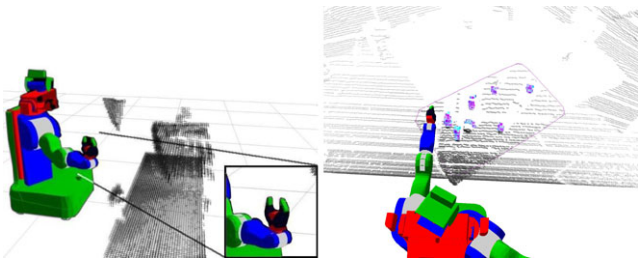

Fig. 7 *Left*: the dynamic obstacle map used for motion planning. *Right*: The extraction of a table surface and the individual object clusters supported by it, from a partial view

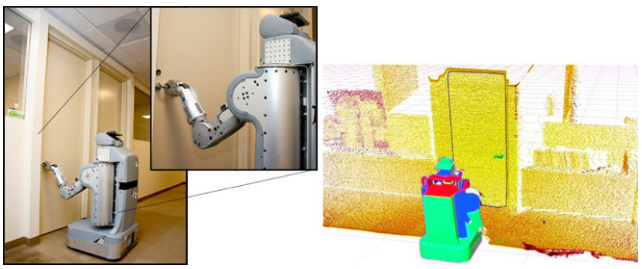

Fig. 8 (Color online) *Left*: Snapshot of the PR2 mobile robot during door identification and opening experiments. *Right*: Door and handle identification example in a 3D point cloud acquired using the tilting Hokuyo laser on the PR2. The handle is marked with *green*, the door frame with *black*, while the rest of the point cloud is shown in intensity (*yellow-red shades*)

The second part of the thesis tackles the semantic scene interpretation of indoor environments, including both methods based on machine learning classifiers and parametric shape model fitting, to decompose the environment into meaningful semantic objects useful for mobile manipulation scenarios (see Fig. 3) [5]. The former is formulated in the context of supervised learning, where multiple training examples are used to estimate the required optimal parameters for the classification of 3D points based on the underlying surface that they represent. The latter includes hierarchical non-linear geometric shape fitting using robust estimators and spatial decomposition techniques, capable of decomposing objects into primitive 3D geometries and creating hybrid shape-surface object models. To obtain informative classes

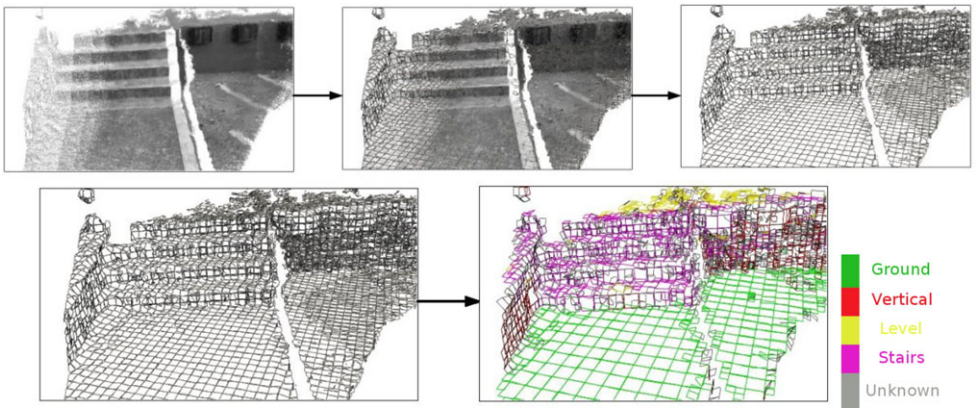

Fig. 9 (Color online) *Top*: Point Cloud model simplification using polygonal fitting. *From left to right*: overlapping aligned point cloud views, combination point cloud data model and polygonal model, and the resulting polygonal model. *Bottom*: Semantic labeling of polygonal surfaces based on heuristic geometrical rules. The labeled classes are: flat and level terrain (*green* and *yellow*), vertical structures (*red*), stairs (*purple*), and unclassified or unknown (*gray*)

of complete objects, a novel global feature estimator which captures the geometry of the object using pairwise relationships between its surface components is proposed and validated on datasets acquired using active stereo cameras.

Finally, the third part of the thesis presents applications of comprehensive 3D perception systems on complete robotic platforms. The first application [7] relates to the problem of cleaning tables by moving the objects present on them in the context of dynamic environments with personal robotic assistants. The proposed architecture constructs 3D dynamic collision maps, annotates the surrounding world with semantic labels, and extracts object clusters supported by tables with real-time performance (see Fig. 7). The experiments are validated on the PR2 (Personal Robotics 2) platform from Willow Garage, using the ROS (Robot Operating System) paradigms (see Fig. 8).

The second application uses the same mobile manipulation platform from Willow Garage, and presents an architecture for door and handle identification from noisy 3D point cloud maps [6]. Experimental results show good robustness in the presence of large variations in the data, without suffering from the classical under or over-fitting problems usually associated with similar initiatives based on machine learning classifiers.

The third and final complete application example tackles the problem of real-time semantic mapping from stereo data for navigation using the RHex (Robot Hexapod) mobile robot from Boston Dynamics [8]. Figure 9 presents results obtained using the system in the SRI campus.

References

1. Rusu RB (2009) Semantic 3d object maps for everyday manipulation in human living environments. PhD thesis, Computer Science department, Technische Universität Müchen, Germany
2. Rusu RB, Marton ZC, Blodow N, Dolha M, Beetz M (2008) Towards 3D point cloud based object maps for household environments. Robot Auton Syst J (Special Issue on Semantic Knowledge)
3. Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (FPFH) for 3D registration. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), Kobe, Japan
4. Rusu RB, Holzbach A, Diankov R, Bradski G, Beetz M (2009) Perception for mobile manipulation and grasping using active stereo. In: Proceedings of the 9th IEEE-RAS international conference on humanoid robots (Humanoids), Paris, France
5. Rusu RB, Marton ZC, Blodow N, Holzbach A, Beetz M (2009) Model-based and learned semantic object labeling in 3D point cloud maps of kitchen environments. In: Proceedings of the 22nd IEEE/RSJ international conference on intelligent robots and systems (IROS), St. Louis, MO, USA
6. Rusu RB, Meeussen W, Chitta S, Beetz M (2009) Laser-based perception for door and handle identification. In: Proceedings of the international conference on advanced robotics (ICAR), Munich, Germany, best paper award
7. Rusu RB, Sucan IA, Gerkey B, Chitta S, Beetz M, Kavraki LE (2009) Real-time perception-guided motion planning for a personal robot. In: Proceedings of the 22nd IEEE/RSJ international conference on intelligent robots and systems (IROS), St. Louis, MO, USA
8. Rusu RB, Sundaresan A, Morisset B, Hauser K, Agrawal M, Latombe JC, Beetz M (2009) Leaving flatland: efficient real-time 3D navigation. J Field Robot